Delft University of Technology

Delft Center for Systems and Control

Technical report 22-017

Hierarchical model predictive control for on-line high-speed railway delay management and train control in a dynamic operations environment*

Y. Wang, S. Zhu, S. Li, L. Yang, and B. De Schutter

Hierarchical model predictive control for on-line high-speed railway delay management and train control in a dynamic operations environment

Yihui Wang, Songwei Zhu, Shukai Li*, Lixing Yang, Bart De Schutter, *Fellow, IEEE*

Abstract—In practice, the operation of high-speed trains is often affected by adverse weather conditions or equipment failures, which result in delays and even cancellations of train services. In this paper, a novel two-layer hierarchical model predictive control (MPC) model is proposed for on-line high-speed railway delay management and train control for minimizing train delays and cancellations. The upper layer manages the global objectives of the train operation, i.e., minimizing the total train delays and providing guidance for the speed control in the lower layer. The objectives of the lower layer are to satisfy the running time requirements given by the upper layer and to save energy at the same time. The optimization problems in both levels of the hierarchical MPC framework are formulated as small-scale mixed integer linear programming problems, which can be solved efficiently by existing solvers. Particularly, the train control problem is solved in a distributed way for each train. Simulation analysis based on the real-life data of the Beijing-Shanghai high-speed railway shows that the proposed hierarchical MPC framework can meet the real-time requirements and reduce train delays effectively when compared with widely accepted strategies, e.g., first-scheduled-first-serve and first-come-first-serve. Moreover, the proposed hierarchical MPC framework also provides good robustness performance for different disturbance scenarios.

Index Terms—high speed railway, train rescheduling, speed control, hierarchical model predictive control, MILP

I. INTRODUCTION

High-speed railway lines are of crucial importance for the mobility of passengers as well as for the competitiveness of regional economy. The reliability and punctuality of high-speed train services are the main focuses of the train operating companies and railway infrastructure managers. In normal situations, high-speed trains are operated as planned in the timetable. However, unavoidable disturbances and disruptions (caused by bad weather, infrastructure failures, malfunction of rolling stocks, etc.) often happen and result in delays and even cancellation of train services. Particularly, in this paper we consider the disturbances that only affect the operation of trains in a local area and the consequences are relatively limited. The disturbances could be caused by adverse weather conditions (e.g., heavy rain and snow, strong wind) or by signaling failures, which generally trigger temporary speed restrictions.

When adverse weather conditions or the equipment failures occur, on the one hand, train dispatchers need to trigger the temporary speed restrictions (TSRs) for the affected area and then take effective dispatching strategies, such as re-timing, re-routing, and re-ordering, for the affected train services to reduce the potential train delays. On the other hand, the automatic train operation (ATO) systems or the train drivers receive the TSRs and the modified departure and arrival (or running and dwell) times via the train control system and they need to take proper driving strategies (e.g., accelerating, decelerating, coasting, and cruising) to respect the instructions given the dispatchers and to save energy consumption. As defined in the literature, the problem faced by the train dispatcher is called the on-line delay management problem or train rescheduling problem, while the problem faced by the ATO systems or train drivers is called the train control problem or the speed control problem. Both the delay management problem and train control problem are well addressed separately in the literature (see the recent survey papers, e.g., [1]–[7]. However, the traffic-related characteristics (i.e., train departure and arrival times, train orders) and the train-related characteristics (i.e., speed trajectories, traction and braking forces) are closely interacted with each other. So we only focus on the studies that elaborate the interaction or integration of traffic management and train control in this paper.

It is common to let the traffic management and train control interact in a sequential way, i.e., the train rescheduling problem is solved first and then the speed profiles are adjusted accordingly. Albrecht [8] presented a sequential approach to investigate the interaction between train rescheduling and train control and concluded that anticipating train control can reduce energy consumption and even reduce delays significantly for certain situations. D'Ariano et al. [9] proposed target points (i.e., arrival times and advisory speeds at key locations) for the speed optimization of rescheduled trains with the aim of punctuality and energy saving. Corman et al. [10] proposed a green wave policy for traffic management, where there is a speed profile available between any two consecutive stations and the dwell times are adjusted to avoid modifications of this speed profile. Caimi et al. [11] proposed a model predictive control approach for rescheduling trains in the complex station area, where there are multiple pre-calculated speed profiles available to build the blocking diagrams for train rescheduling.

Another way to coordinate the traffic management and train control is using iterative approaches, where speed profiles can be optimized according to the updated train timetable and then be fed back to the train rescheduling process for the sake of performance enhancement. Mazzarello and Ottaviani [12] developed a double feedback loop structure to optimize the train-related and traffic-related decisions using heuristics. Similarly, an iterative train rescheduling and speed adjusting approach is presented in [13], where the drivers adjust the speed of trains in order to meet the updated train timetable. Moreover, [14] investigated the multiple-phase train trajectory optimization approach with consideration of constraints imposed by traffic management.

Unified models for the integration of traffic management and speed control approach have become popular in recent years. Yin et al. [15] formulated a train rescheduling model for metro lines that considers the time-variant passenger dynamic characteristics and the energy consumption of train operations. Xu et al. [16] presented a mixed integer linear programming model based on alternative graphs for the integration of traffic management and speed control, where the operation speeds of trains are classified into several levels and managed by indicating additional travel times for the train rescheduling process. [17] proposed an integrated optimization model for traffic management and train control, where the train speed profile is computed to determine the main target points (i.e., target position, target time, target speed) and sub-target points for preventing potential traffic conflicts. Luan et al. [18], [19] studied the integration problem of train management and speed control by developing three optimization models to form the real-time train schedule via optimizing the driving strategies. However, the integrated model is very complex and many approximations and simplifications are needed to obtain solutions efficiently.

The scale of the integrated traffic management and train control problem is too large for real-time application in most of the cases, which becomes computationally prohibitive for on-line decision-making in a dynamic environment. Moreover, the previous literature on the integrated optimization of train traffic management and train control is mainly confined to static optimization, which seldom involves dynamic updated information [18], [15], [16]. In addition, to realize the on-line decision making process for the traffic management, the model predictive control (MPC) algorithm has been developed with the strategy of rolling optimization [20]–[23], which can be efficiently applied to cope with the on-line high-speed railway delay management and train control problem.

In order to satisfy the on-line requirements of practical applications, we therefore present a hierarchical MPC approach for the on-line delay management and train control of the high-speed railway. The contributions of the current paper to the literature are listed as follows:

- A hierarchical optimization framework is presented for the on-line high-speed railway operations in a dynamic operations environment, where the delay management problem is addressed at a slow time scale in the upper

layer, while the speed control problem is implemented in a fast time scale in the lower layer.
- A novel hierarchical model predictive control algorithm is particularly designed for the on-line delay management and train control problem, which is solved in a moving-horizon manner, where the real-time updated information of the train, even considering uncertain operational conditions, can be included in the on-line decision-making process. This makes the solutions robust to uncertainty, disturbances, and even model mismatch.
- The hierarchical model predictive control algorithm divides the original problem into two-layer small-scale optimization problems that can be solved relatively independently to realize the complexity reduction. More specifically, for the train control problem in the lower layer, the speed profile problem for each train can be solved in parallel, which greatly improves the computational efficiency to meet the on-line calculation requirements of high-speed trains.

The remainder of this paper is organized as follows. In Section II, we give a problem statement and the formulation assumptions. Section III introduces the mathematical formulation of the hierarchical model predictive control scheme. The solution approach for the hierarchical MPC framework is proposed in Section IV. Experimental results based on the Beijing-Shanghai high-speed railway line are given in Section V. Finally, Section VI concludes the paper.

II. PROBLEM STATEMENT AND FORMULATION ASSUMPTIONS

In this section, we first introduce the topology and the signaling principle of a high-speed railway line. The current practice for the handling process of the temporary speed restrictions (TSRs) is then described. After that we illustrate the hierarchical model prediction control framework by a small example. Finally, the assumptions of this research are listed and explained in detail.

A. Problem statement

In practical operations, unavoidable perturbations (or disturbances) caused by adverse weather conditions or equipment failures often happen and significantly impact the operation of trains. In order to cope with unavoidable perturbations (or disturbances), temporary speed restrictions (TSRs) need to be triggered by the dispatchers and put into the signaling systems to address safety concerns. However, there are also more serious cases that are classified as "disruptions" where tracks may be totally blocked and some train services need to be suspended and fully stopped. In this paper, we only consider the disturbances or perturbations that trigger TSRs and that only affect trains in a relatively small area. In practice, the starting and end positions of the TSRs are also at the splitting points between block sections as illustrated by the yellow and orange areas in Figure 1.

Moreover, when TSRs are triggered due to adverse weather conditions or equipment failures, train drivers (or automatic train operation (ATO) systems) of the affected trains should

apply braking forces to reduce the speed of the trains to satisfy the speed limits. If drivers or ATO systems cannot reduce the speed of the trains in time, emergency braking may be triggered by the signaling system to stop trains immediately. So it is important to have a speed control model that takes the detailed train characteristics, fixed speed limits, TSRs, arrival times, etc. into account to make sure that trains arrive at the next station smoothly and save energy consumption as well. Since the affected trains are operated at a lower speed, which results in longer running time between stations, there could be potential conflicts for the operation of neighboring trains. So a train rescheduling model is required in the control center to take proper measures, e.g., re-timing (changing running and dwell times of trains) and re-ordering (changing the order of trains), for resolving the conflicts and minimizing delays from the planned timetable.
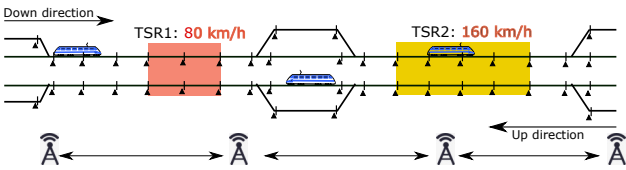

Fig. 1: The illustration of the track layout and temporary speed restrictions (TSRs) of a high-speed railway line (the transponders (triangles) and the signal towers are used to transmit TSRs information and a segment corresponds to a block section)
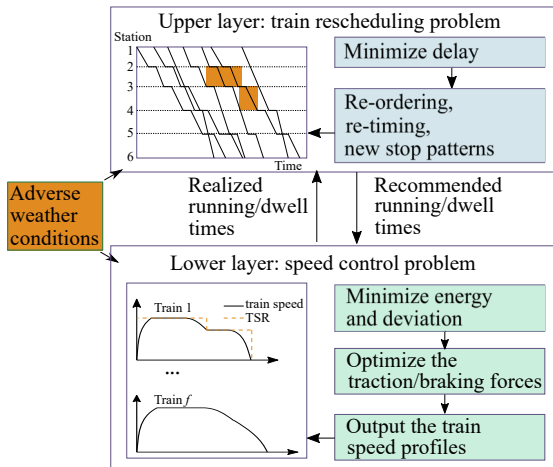

Fig. 2: The hierarchical traffic management and train control framework for high speed railways

In this paper, we propose a hierarchical model predictive control framework, where the problems in the upper and lower levels are both solved in a rolling horizon way. The current train rescheduling and train control framework in practice is a hierarchical setting basically. The dispatchers in the control center regulate the order of trains, the running/dwelling times, etc, according to the measured departure and arrival times of trains at stations and the updated TSRs information. The train drivers or ATO systems adapt the behavior (dwelling, speed, etc.) of trains according to the measured statues (position,

speed, etc.) of itself and the trains nearby and the running/dwelling times and TSRs specified by the dispatchers. In this paper, we propose a hierarchical model predictive control framework as illustrated in Figure 2 for the traffic management and train control of high speed railways under unavoidable perturbations (or disturbances), where TSRs are triggered to ensure the safe operation of trains. Within the hierarchical model predictive control framework, the original integrated train rescheduling and speed control problem is effectively reduced and decomposed into two-layer optimization problem. The information that are transmitted between the upper layer and the lower layer involves the scheduled running/dwell times and the realized running/dwell times for trains, as illustrated in Figure 2. Since the disturbances cannot be fully avoided by the anticipation and the train schedule obtained in the high level could involve potential conflicts, the feedback of the upper layer is the actual departure/arrival times that yield in the lower level controller. Moreover, the higher level controller are triggered when the actual departure/arrival delay reaches a pre-defined threshold or the information of the TSRs is updated. The rolling horizon could be expressed as a certain time span (e.g., an hour) or a certain number of trains (e.g., 5 trains). The lower level updates the speed profiles with a much higher frequency, which updates when a TSR is triggered, a train departs or arrives at a station, and/or a train passes a block section. Compared with the rolling horizon in the upper level, the length of the rolling horizon in the lower level is much shorter and it could e.g. be equal to the left running time to the next station. Therefore, the hierarchical model predictive control framework does not only reduce the computational complexity of the integrated traffic management and train control problem, but also provides a on-line decision making approach with the real-time updated information in a dynamic operations environment.

B. Assumptions

In our hierarchical model predictive control framework, we make the following assumptions:

- The operation of train services in the up and down directions is separated from each other; so we can consider the operation of trains in one direction only.
- The cancellation of train services is not considered in the train rescheduling problem since we only consider small disturbances and perturbations caused by adverse weather conditions or signal failures (e.g., short communication interruptions).
- The rolling stock circulation plan is not considered in the train rescheduling formulation.

III. MATHEMATICAL FORMULATION

Within the hierarchical optimization framework, two optimization models are introduced: (1) a macroscopic train rescheduling model in the upper layer considering a high-speed line with multiple stations and different types of train services, and (2) a speed control model in the lower layer with consideration of the detailed train characteristics, line conditions, temporary speed limits, etc. In this section, the

TABLE I: Parameters and subscripts for the model formulation

Notations	Definition
$\mathbb{I}, \mathbb{F}$	set of stations and trains
i, f	indices of stations and trains
$\mathbb{I}_f, \mathbb{I}_f^{\text{stop}}$	set of stations that train f passes and stops at
I_f^{stop}	number of scheduled stops of train f
$\mathbb{F}_i$	set of trains pass station i
$\mathbb{F}_{i,i+1}$	set of trains pass station i and station $i+1$
$\bar{d}_{f,i}, \bar{a}_{f,i}$	planned departure time and arrival time of train f at station i
$\bar{w}_{f,i}$	planned dwell time for train f at station i
$\sigma_{f,i}^{\text{r}}, \sigma_{f,i}^{\text{w}}$	disturbances of running and dwell time of train f at station i
$h_{f,f'}^{\text{min,arr}}, h_{f,f'}^{\text{min,dep}}$	minimum arrival-arrival and departure-departure headways
C_i	capacity of station i
$N_{i,i'}$	total number of small intervals between station i and i'
n	index of the small intervals between station i and i'
$E_{i,n}^{\text{max}}$	maximal kinetic energy per mass for a train at position $s_{i,n}$
E_{min}	minimum kinetic energy (a small positive value) per mass for a train
ℓ	index of temporary speed restrictions
$\mathbb{L}$	set of temporary speed restrictions
E_ℓ^{TSR}	maximal kinetic energy of the ℓ-th temporary speed restriction
$\underline{i}_\ell, \bar{i}_\ell$	start and end stations of the ℓ-th temporary speed restriction
$T_\ell^{\text{start}}, T_\ell^{\text{end}}$	start and end time of the ℓ-th temporary speed restriction

TABLE II: Variables for the model formulation

Notations	Definition
$d_{f,i}, a_{f,i}$	departure and arrival time of train f at station i
$r_{f,i}, w_{f,i}$	running and dwell time of train f at station i
$Y_{f,i,\ell}$	binary variable to indicate whether the operation of train f between i and $i+1$ is affected by the ℓ-th TSR
$\xi_{f,f',i}$	binary variable to indicate the arrival order of train f and f' at station i
$\eta_{f,f',i}$	binary variable to indicate the departure order of train f and f' at station i
$\varepsilon_{f,f',i}$	binary variable to indicate the order of train f's arrival and train f''s departure
$a_{f,i,n}, d_{f,i,n}$	arrival and departure time of train f at position $s_{i,n}$ between stations i and i'
$E_{f,i,n}$	kenetic energy of train f at position $s_{i,n}$ between stations i and i'
$u_{f,i,n}$	traction/braking force of train f at position $s_{i,n}$ between stations i and i'

notations and decision variables are introduced first. Then, the constraints and objective functions of the train rescheduling and speed profile optimization are formulated.

A. Notations and decision variables

Table I lists the parameters and subscripts used in the model formulation. Moreover, Table II gives the decision variables for the railway traffic delay management and train control.

B. Upper hierarchical layer

We consider a high speed railway line with I stations, i.e., $i \in \mathbb{I} = \{1, 2, \ldots, I\}$, where station $i+1$ is the first station after station i and the train services that operate on this line are indexed by f with $f \in \mathbb{F} = \{1, 2, \ldots, F\}$. The set of stations that train f passes is denoted by $\mathbb{I}_f$ and the set of stations in which train f should stop is defined as $\mathbb{I}_f^{\text{stop}}$. We have $\mathbb{I}_f \subseteq \mathbb{I}$ and $\mathbb{I}_f^{\text{stop}} \subseteq \mathbb{I}_f$. In addition, the number of scheduled stops for train service f is denoted as I_f^{stop}.

The objective function of the train rescheduling model is to minimize the sum of the mean absolute delay time at all scheduled stops for all trains, i.e., the deviations of the arrival and departure times with respect to the planned arrival times:

$$\min \quad Z_{\text{up}} = \sum_{f \in \mathbb{F}} \sum_{i \in \mathbb{I}_f^{\text{stop}}} \left(\frac{|a_{f,i} - \bar{a}_{f,i}|}{I_f^{\text{stop}}} + \frac{|d_{f,i} - \bar{d}_{f,i}|}{I_f^{\text{stop}}} \right), \quad (1)$$

where $a_{f,i}$ ($d_{f,i}$) and $\bar{a}_{f,i}$ ($\bar{d}_{f,i}$) are the actual and planned arrival (departure) times of train f at station i.

The constraints for the train rescheduling problem in the upper layer involves the following five groups.

1) Departure/arrival constraints: The departure and arrival times of train services should satisfy the following constraints

$$a_{f,i+1} = d_{f,i} + r_{f,i} + \sigma_{f,i}^{\text{r}}, \quad (2)$$

$$d_{f,i} = a_{f,i} + w_{f,i} + \sigma_{f,i}^{\text{w}}, \quad (3)$$

where $r_{f,i}$ is the running time of train f between station i and station $i+1$, which depends on the train categories and characteristics of train f, line conditions between station i and $i+1$, stop patterns, temporary speed restrictions, etc., and $w_{f,i}$ is the dwell time of train f at station i. Moreover, $\sigma_{f,i}^{\text{r}}$ and $\sigma_{f,i}^{\text{w}}$ are introduced to denote the disturbances of the running time and dwell time of train f at station i, which are given based on estimations. In addition, the departure time $d_{f,i}$ should be later than or equal to the planned departure time if train f stops at station i, i.e.,

$$d_{f,i} \geq \bar{d}_{f,i}. \quad (4)$$

2) Running/dwell time constraints: The running time $r_{f,i}$ should be larger than or equal to the minimum running time $r_{f,i}^{\text{min}}$, i.e.,

$$r_{f,i} \geq r_{f,i}^{\text{min}}, \quad (5)$$

where the minimum running times for the normal operations can be calculated based on the speed limits, characteristics of trains, etc. If the operation of train f is affected by a TSR indexed by ℓ, as illustrated in Figure 3, then the minimum running time of train f between station i and $i+1$ will be much longer compared with the one for the normal operation. A binary variable $Y_{f,i,\ell}$ is introduced to indicate whether the operation of train f between stations i and $i+1$ is fully affected by TSR indexed by ℓ, which is defined as follows

$$Y_{f,i,\ell} = \begin{cases} 1 & \text{if } d_{f,i} \geq T_\ell^{\text{start}} \text{ and } a_{f,i+1} \leq T_\ell^{\text{end}}, \ i, i+1 \in \mathbb{I}_\ell, \\ 0 & \text{otherwise,} \end{cases}$$

$$(6)$$

where $\mathbb{I}_\ell$ denotes the set of stations influenced by TSR ℓ and $Y_{f,i,\ell} = 1$ means that the operation of train f between station i and $i+1$ is affected by TSR ℓ, i.e., the running process of train f between stations i and $i+1$ is a subset of the activation period of TSR ℓ. If the operation of train f between station i and $i+1$ is affected by TSR ℓ, the minimum running time $r_{f,i}^{\text{min}}$ cannot be achieved. Hence, an updated minimum running time, i.e., $r_{f,i,\ell}^{\text{min}}$ with $r_{f,i,\ell}^{\text{min}} > r_{f,i}^{\text{min}}$, is adopted to increase the feasibility of the train schedule generated in the upper level. The following constraint is introduced to consider the updated minimum running time, i.e.,

$$r_{f,i} \geq r_{f,i,\ell}^{\text{min}} + M(Y_{f,i,\ell} - 1), \quad (7)$$
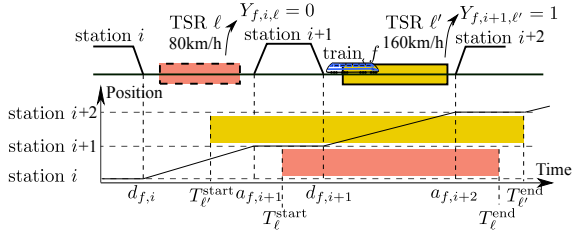
Fig. 3: An illustration of temporary speed restrictions (TSRs) on the operation of train f

Fig. 4: An illustration of the minimum headways between trains

where M is a sufficient large number, which is introduced to ensure $r_{f,i} \geq r_{f,i,\ell}^{\min}$ when binary variable $Y_{f,i,\ell}$ is equal to 1, but to let this constraint be fully satisfied when $Y_{f,i,\ell}$ is equal to 0. For constraint (7), the value of M should be larger than $r_{f,i,\ell}^{\min} - r_{f,i}^{\min}$. It is worth to note that the operation of train f between stations i and $i+1$ could be partially affected by TSRs. Due to the difficulty of accurate estimation of minimum running times, if the operation of train between two consecutive stations is partially affected by TSRs, only the minimum running time constraint (5) is involved. This could result in potential train conflicts; however, these conflicts can be well taken care of in the lower level control.

It is worth to note that the maximum running times are not specified in the upper layer model because the values of the TSRs can not be known in advance and it decides the maximum running times. However, the delay of train services are considered in the objective function, which indirectly considers the maximum running times.

Moreover, if train service f does not stop at station i, then the dwell time $w_{f,i}$ should be equal to zero. In addition, if a train is scheduled to stop at a certain station, then this station cannot be skipped during the rescheduling so as to guarantee the satisfaction of passengers. Therefore, in our paper, we only allow trains to add new stops if needed. Therefore, we have the following constraints for the dwell times

$$w_{f,i} \geq \bar{w}_{f,i}, \qquad (8)$$

where $\bar{w}_{f,i}$ is equal to zero if train f does not plan to stop at station i. So when we have $w_{f,i} > \bar{w}_{f,i}$ and $\bar{w}_{f,i} = 0$, it means that the stop pattern of train f is changed at station i.

3) Headway constraints: The operation of train services should also satisfy the headway constraints, which involve the arrival-arrival headway constraints, the departure-departure headway constraints, the arrival-departure headway constraints, and the departure-arrival headway constraints. Since we only consider the macroscopic train scheduling model in this paper and the detailed station layout is not considered, we assume that the departure routes and the arrival routes of trains at stations are independent from each other, which is generally holds for most of the high speed railway stations in China. So we only consider the arrival-arrival and departure-departure headway constraints as illustrated in Figure 4, which can be formulated as follows:

$$a_{f,i} - a_{f',i} \geq h_{f,f'}^{\min,\text{arr}} - M\xi_{f,f',i}, \qquad (9)$$

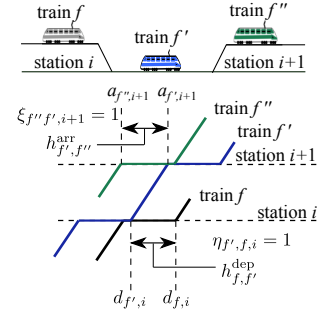$$d_{f,i} - d_{f',i} \geq h_{f,f'}^{\min,\text{dep}} - M\eta_{f,f',i}, \qquad (10)$$

where $\xi_{f,f',i}$ and $\eta_{f,f',i}$ denote the arriving and departing order of train f and f' at station i, i.e.,

$$\xi_{f,f',i} = \begin{cases} 1 & \text{if train } f \text{ arrives earlier than train } f', \\ 0 & \text{if train } f \text{ arrives later than train } f', \end{cases} \qquad (11)$$

and

$$\eta_{f,f',i} = \begin{cases} 1 & \text{if train } f \text{ departs earlier than train } f', \\ 0 & \text{if train } f \text{ departs later than train } f'. \end{cases} \qquad (12)$$

It is worth to note that the big M in (9) and (10) should be larger than $h_{f,f'}^{\min,\text{arr}} - a_{f,i}^{\min} + a_{f',i}^{\max}$ and $h_{f,f'}^{\min,\text{dep}} - d_{f,i}^{\min} + d_{f',i}^{\max}$.

4) Train ordering constraints: The order of trains can only be changed at stations in our paper, so we have

$$\xi_{f,f',i+1} = \eta_{f,f',i}, \qquad (13)$$

which means that the arrival order of trains f and f' at station $i+1$ should be the same as the departure order at station i because there is no overtaking facilities between two stations. Furthermore, the order between any two trains should also satisfy

$$\xi_{f,f',i} + \xi_{f',f,i} = 1, \forall f \in \mathbb{F}, f' \in \mathbb{F}, i \in \mathbb{I}_f, i \in \mathbb{I}_{f'}, \qquad (14)$$

$$\eta_{f,f',i} + \eta_{f',f,i} = 1, \forall f \in \mathbb{F}, f' \in \mathbb{F}, i \in \mathbb{I}_f, i \in \mathbb{I}_{f'}, \qquad (15)$$

which indicates that either train f arrives at station i before train f' or train f' arrives before train f at station i.

5) Station capacity constraints: Each station has a capacity limit, which corresponds to the maximum number of trains that can stop in or pass a station simultaneously. According to the practical operation rules, a station track (platform or siding) can only accommodate one train at most and the station tracks are usually dedicated to one operation direction, i.e., either the up direction or the down direction. So when considering the station capacities, we could only take the trains operated in the same direction into account. We introduce binary variables $\varepsilon_{f,f',i}$ to indicate whether the arrival of train f is before the departure of train f' or not, i.e.,

$$\varepsilon_{f,f',i} = \begin{cases} 1 & \text{if train } f \text{ arrives before the departure of } f', \\ 0 & \text{if train } f \text{ arrives after the departure of } f'. \end{cases} \qquad (16)$$

The station capacity constraints can be formulated as

$$\sum_{f' \in \mathbb{F}, i \in \mathbb{I}_{f'}} (1 - \xi_{f,f',i}) - \sum_{f' \in \mathbb{F}, i \in \mathbb{I}_{f'}} (1 - \varepsilon_{f,f',i}) \leq C_i - 1, \qquad (17)$$

for all $f \in \mathbb{F}, i \in \mathbb{I}_f$, where C_i is the capacity[1] of station i. More specifically, $\sum_{f' \in \mathbb{F}, i \in \mathbb{I}_f}(1 - \xi_{f,f',i})$ denotes the number of trains that have arrived at station i before the arrival of train f and $\sum_{f' \in \mathbb{F}, i \in \mathbb{I}_f}(1 - \varepsilon_{f,f',i})$ denotes the number of trains that have departed from station i before the arrival of train f. In order to let train f enter station i, there should be a station track available. So the right-hand side of inequality (17) is $C_i - 1$.

The assembled optimization problem for the upper layer can be written as follows:

$$\min_{(a_{f,i}, d_{f,i}, r_{f,i}, \xi_{f,f',i}, \eta_{f,f',i})} \sum_{f \in F} \sum_{i \in I} \frac{|a_{f,i} - \bar{a}_{f,i}|}{I_F^{\text{stop}}} + \frac{|d_{f,i} - \bar{d}_{f,i}|}{I_F^{\text{stop}}} \quad (18)$$
$$\text{s.t.} \quad (2) - (17).$$

C. Lower hierarchical layer

The upper layer specifies the running time $r_{f,i}$ of train f between station i and $i+1$ by using the macroscopic train rescheduling model to minimize the deviation of train operations with respect to the planned timetable. However, train f may not stop at all stations, i.e., train f may skip some stations according to the planned timetable. For the train control problem, the speed profile is optimized between two consecutive stations where train f stops, denoted by stations i and i'. Therefore, $\bar{r}_{f,i,i'}$ is introduced to denote the running time of train f between stations i and i' that are recommended by the upper layer, which can be calculated by $\bar{r}_{f,i,i'} = \sum_{\tilde{i}=i}^{i'-1} r_{f,\tilde{i}}$. The detailed information, such as the characteristics of trains, the section length, the grade profile and curvature of the track between those two stations, is not explicitly considered in the upper level model, which only takes the minimum running times between stations into account. So the running times calculated by the upper layer may not be feasible for the detailed train operation model in the lower level, especially if there are some TSRs appear due to the strong wind, heavy rain, big snow, etc. Those TSRs would largely affect the running time of trains between stations. For example, the maximum speed for the type G trains is 350 km/h in CTCS-3, however, the TSR could be 200 km/h, 120 km/h, or even 60 km/h for a certain time period in the affected block sections. Therefore, the specified running times in the upper layer may not be realized and delays would occur and propagate among the trains in the high speed lines.

The mass-point model is usually used in the literature and the operation of trains is subject to Newton's second law. As presented in previous studies [24]–[27], it is better to select the position s as an independent variable instead of the time t because the speed limits, grade profiles, curvatures vary with respect to the position of trains. Furthermore, as in [25], [26], [28], we take kinetic energy per mass unit $\tilde{E} = 0.5v^2$ and time t as states to eliminate some (but not all) of the model nonlinearities. As given in [28], the continuous-space train dynamic model can be written as

$$m\rho \frac{\mathrm{d}\tilde{E}}{\mathrm{d}s} = u(s) - R_{\text{b}}(\tilde{E}) - R_{\text{l}}(s, \tilde{E}), \quad (19)$$
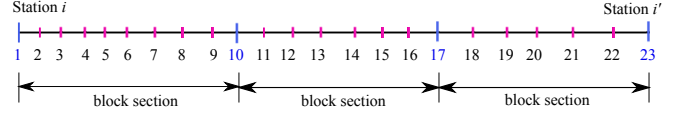
Fig. 5: An illustrative example for the splitting of block sections between two stations

$$\frac{\mathrm{d}t}{\mathrm{d}s} = \frac{1}{\sqrt{2\tilde{E}}}, \quad (20)$$

where m $(m > 0)$ is the train's mass [kg], ρ $(\rho > 0)$ is the rotating mass factor [28], v $(v \geq 0)$ is the train's speed [m/s], s $(s \geq 0)$ is the train's position [m], u corresponds to the traction or braking force [N], $R_{\text{b}} : \mathbb{R} \to \mathbb{R}$ is the basic resistance including roll resistance and air resistance [N], and $R_{\text{l}} : \mathbb{R} \to \mathbb{R}$ is the line resistance caused by track grade, curves, and tunnels [N]. More specifically, the basic resistance and the line resistance can be calculated by

$$R_{\text{b}}(\tilde{E}) = R_{\text{b},1} + 2R_{\text{b},2}\tilde{E}, \quad (21)$$
$$R_{\text{l}}(s, \tilde{E}) = \tilde{R}^{\text{l}}(s) + 2\tilde{R}^{\text{t}}(s)\tilde{E}, \quad (22)$$

where $R_{\text{b},1}$ and $R_{\text{b},2}$ are the coefficients of the basic resistance, $\tilde{R}^{\text{l}}(s)$ and $\tilde{R}^{\text{t}}(s)$ are the components of line resistance that do not depend on the train's speed.

In order to formulate the microscopic train operation control model, the block sections between stations are considered, where the length of a block section is around 1000 m to 1500 m. Each block section is then split into small intervals with length of e.g. 50 m or 100 m for the detailed computation of speed profiles. The position of the nodes that divided those small intervals between stations i and i' are denoted by $s_{i,0}, s_{i,1}, \ldots, s_{i,n}, \ldots, s_{i,N_{i,i'}}$ and the length of those intervals can be calculated by $\Delta s_{i,n} = s_{i,n} - s_{i,n-1}$ for $n \in \mathbb{N}_{i,i'} = \{1, 2, \ldots, N_{i,i'}\}$. The splitting points for the block sections are given in $\mathbb{N}_{i,i'}^{\text{B}}$. An illustrative example for the small intervals is given in Figure 5, where there are three block sections between stations i and i' and each block section is then split into multiple small intervals. Particularly, the splitting points of the block sections are 1, 10, 17, and 23, i.e., $\mathbb{N}_{i,i'}^{\text{B}} = \{1, 10, 17, 23\}$ for the small example given in Figure 5.

Particularly, $s_{i',0}$ is basically the position of station i' and we have $s_{i,N_{i,i'}} = s_{i',0}$ if station i' is the next station that train f stops at after station i. By appropriately defining the discretization of the intervals, without loss of generality the coefficient of the line resistance can be written as follows

$$\tilde{R}^{\text{l}}(s) = R_{i,n}^{\text{l}}, \text{ for } s \in [s_{i,n-1}, s_{i,n}] \text{ and } n \in \mathbb{N}_{i,i'}, \quad (23)$$
$$\tilde{R}^{\text{t}}(s) = R_{i,n}^{\text{t}}, \text{ for } s \in [s_{i,n-1}, s_{i,n}] \text{ and } n \in \mathbb{N}_{i,i'}. \quad (24)$$

In addition, we introduce $a_{f,i,n}$ and $d_{f,i,n}$ to denote the arrival and departure times of train f at node (i, n) between stations i and i'. We then have

$$a_{f,i,0} = d_{f,i}, \forall f \in \mathbb{F}, i \in \mathbb{I}_f, \quad (25)$$
$$d_{f,i,N_{i,i'}} = a_{f,i'}, \forall f \in \mathbb{F}, i, i' \in \mathbb{I}_f, \quad (26)$$

which means that when train f departs from station i, it immediately arrives at node 0 between stations i and i' and train f arrives at station i immediately when it departs from

from the last node $N_{i,i'}$ between station i and i'. When the arrival time $a_{f,i}$ is not equal to $d_{f,i}$, this means that train f dwells at station i. In addition, we have

$$d_{f,i,n} = a_{f,i,n}, \forall f \in \mathbb{F}, i \in \mathbb{I}, n \in \mathbb{N}_{i,i'}/\{N_{i,i'}\}, \quad (27)$$

which means that train f is not allowed to dwell at the intermediate intervals between stations i and i'. By solving the differential equations (19) and (20), we obtain

$$E_{f,i,n} = \alpha_{f,i,n}E_{f,i,n-1} + \beta_{f,i,n}u_{f,i,n} + \gamma_{f,i,n}, \quad (28)$$

$$a_{f,i,n} = d_{f,i,n-1} + \frac{1}{2}\left(\frac{1}{\sqrt{2E_{f,i,n-1}}} + \frac{1}{\sqrt{2E_{f,i,n}}}\right)\Delta s_{i,n}, \quad (29)$$

for all $f \in \mathbb{F}, i \in \mathbb{I}, n \in \mathbb{N}_{i,i'}$, where $\alpha_{f,i,n} = e^{-2\Delta s_{i,n}(R_{b,2}+\tilde{R}_{i,n}^t)/m\rho}$, $\beta_{f,i,n} = -\frac{1}{2(R_{b,2}+\tilde{R}_{i,n}^t)}\left(e^{-2\Delta s_{i,n}(R_{b,2}+\tilde{R}_{i,n}^t)/m\rho} - 1\right)$, and $\gamma_{f,i,n} = \frac{R_{b,1}+\tilde{R}_{i,n}^l}{2(R_{b,2}+\tilde{R}_{i,n}^t)}\left(e^{-2\Delta s_{i,n}(R_{b,2}+\tilde{R}_{i,n}^t)/m\rho} - 1\right)$. See [28] for more details of these calculations.

The operation of trains should satisfy the constraints introduced by the train dynamics, line characteristics, etc. Specifically, the traction and braking forces $u_{f,i,n}$ should be larger than the maximum braking force (a negative value) and smaller than the maximum traction force, i.e.,

$$u_{\min} \le u_{f,i,n} \le u_{\max}(\sqrt{2E_{f,i,n}}), \forall f \in \mathbb{F}, i, i' \in \mathbb{I}_f, n \in \mathbb{N}_{i,i'}, \quad (30)$$

where the maximum braking force is considered as a constant and the maximum traction force is a nonlinear function of the train's speed[2] as in [28]. It is worth to note that the traction/braking force $u_{f,i,n}$ is a constant for small interval $[s_{i,n-1}, s_{i,n}]$. So the operation regime of trains (e.g., acceleration, deceleration, coasting or cruising) in a small interval does not change in this small interval. Therefore, if the speed limits are respected at the nodes of this interval then all the intermediate points inside the interval also satisfy the constraints caused by speed limits, i.e.,

$$0 \le E_{f,i,n} \le E_{i,n}^{\max}, \forall f \in \mathbb{F}, i, i' \in \mathbb{I}_f, n \in \mathbb{N}_{i,i'}. \quad (31)$$

Note that if train f dwells at station i, then the speed of trains should be zero when it stops. However, since we have $E_{f,i,n}$ in the denominator in (29), we introduce a small positive number $E_{\min}$, such that $E_{f,i,n} \ge E_{\min} > 0$. This means that the speed of trains is always strictly larger than zero, which is not restrictive in practice as stated in [26], [28]. Therefore, if train f dwells at station i, we have the following constraints

$$E_{f,i,0} = E_{\min}, \quad (32)$$

$$E_{f,i,N_{i,i'}} = E_{\min}. \quad (33)$$

Furthermore, during the operation process, there could be temporary speed constraints caused by extreme weather or equipment failures, under which a TSR generally appears and affects a certain area (defined by block sections) for a certain time period. In addition, it is possible that there are several

[2]The maximum traction force that could be taken depends on the train speed, which may vary in the interval $[s_{i,n-1}, s_{i,n}]$. Since here we discretize the problem, only the speeds at the nodes are actually taken into account. So the value of the maximum allowable traction force could also be $u_{\max}(\sqrt{2E_{f,i,n+1}})$ or $u_{\max}(\sqrt{E_{f,i,n}+E_{f,i,n+1}})$.

TSRs that need to be respected simultaneously in a high-speed railway line. A TSR is denoted by a tuple

$$\left(E_\ell^{\text{TSR}}, T_\ell^{\text{start}}, T_\ell^{\text{end}}, s_{\underline{i}_\ell, \underline{n}_\ell}, s_{\bar{i}_\ell, \bar{n}_\ell}\right) \quad (34)$$

where E_ℓ^{TSR} indicates the maximum speed that can be operated by trains, T_ℓ^{start} and T_ℓ^{end} are the start and end time of this TSR, $s_{\underline{i}_\ell, \underline{n}_\ell}$ and $s_{\bar{i}_\ell, \bar{n}_\ell}$ indicate the start and end positions of the TSR. Note that these two positions correspond to the start and end of block sections, i.e., $\underline{n}_\ell \in \mathbb{N}_{i_\ell, i_\ell'}^{\text{B}}$ and $\bar{n}_\ell \in \mathbb{N}_{i_\ell, i_\ell'}^{\text{B}}$. Here we introduce $\mathbb{N}_\ell^{\text{TSR}}$ to denote all the nodes that are affected by TSR ℓ. In addition, $\mathbb{N}_f$ is introduced to denote all the nodes passed by train f. Therefore, we need to determine whether train f is affected by TSR ℓ or not, i.e.,

$$T_\ell^{\text{start}} \le d_{f,i_\ell,n_\ell} \le T_\ell^{\text{end}}. \quad (35)$$

So if (35) is satisfied, then train f is affected by TSR ℓ and is required to respect the following constraints

$$E_{f,i,n} \le E_\ell^{\text{TSR}}, \forall (i,n) \in \mathbb{N}_\ell^{\text{TSR}} \cap \mathbb{N}_f. \quad (36)$$

Moreover, the safe operation of train f is guaranteed by the advanced signaling systems as mentioned in Section II-A. When TSRs are triggered, the operation of trains may be highly affected by the signaling system. Therefore, it is important to consider the interaction between neighboring trains to save energy by avoiding unnecessary acceleration and deceleration. In the lower layer of our hierarchical framework, the optimal control problem of each train considers the behavior, i.e., the speed profile, of its preceding train. Let train f' be the predecessor of train f. Here, we adopt the approach proposed by [16] to describe the speed limits introduced by the preceding train f', where the speed of train f is dependent on the number of free block sections $n_{f,i,n}^{\text{b}}$ between train f and train f' when train f is at position $s_{i,n}$, i.e.,

$$\begin{cases} \sqrt{2E_{f,i,n}} \le 300 \text{ km/h}, & \text{if } n_{f,i,n}^{\text{b}} = 5, \\ \sqrt{2E_{f,i,n}} \le 250 \text{ km/h}, & \text{if } n_{f,i,n}^{\text{b}} = 4, \\ \sqrt{2E_{f,i,n}} \le 200 \text{ km/h}, & \text{if } n_{f,i,n}^{\text{b}} = 3, \\ \sqrt{2E_{f,i,n}} \le 160 \text{ km/h}, & \text{if } n_{f,i,n}^{\text{b}} = 2, \\ \sqrt{2E_{f,i,n}} \le 120 \text{ km/h}, & \text{if } n_{f,i,n}^{\text{b}} = 1. \end{cases} \quad (37)$$

The value of $n_{f,i,n}^{\text{b}}$ is determined by the arrival and departure times of train f and its preceding train f'. Particularly, the arrival and departure times of train f' are based on its speed profile obtained in the previous time step because the train control problem are solved for each train independently in the lower level. We introduce $\text{pred}(k, n_{\text{b}})$ to denote the index of the block section that is n_{b} block sections preceding the block section that involves position $s_{i,n}$. If $a_{f,i,n} \ge d_{f',i,\text{pred}(n,5)}$, i.e., there are more than five free block sections between trains f and f', the operation of train f' does not affect train f. However, if $d_{f',i,\text{pred}(n,4)} \le a_{f,i,n} \le d_{f',i,\text{pred}(n,5)}$, then we have $n_{f,i,n}^{\text{b}} = 5$, which means that the speed of train f at $s_{i,n}$ should be less than 300 km/h. The other speed restriction constraints caused by the preceding train can be introduced in a similar way.

In order to ensure the feasibility of the train operation problem in the lower layer, the running time of train f

between two dwelling stations, denoted by i and i', cannot be formulated as hard constraint, but should be included as a soft constraint in the objective function. Additionally, the total energy consumption of the trains is also considered in the objective function. Thus, the objective function can be given as follows

$$\min \quad Z_{\text{low},f,i,i'} = \sum_{(\tilde{i},k)\in\mathbb{N}_{f,i,i'}} \theta_1 \max(u_{f,\tilde{i},k},0)\Delta s_{\tilde{i},k} + \theta_2\left|r_{f,i,i'} - \bar{r}_{f,i,i'}\right|, \tag{38}$$

where the set of all the nodes between station i and i' that are passed by train f is denoted as $\mathbb{N}_{f,i,i'}$, $r_{f,i,i'}$, i.e., $a_{f,i'} - d_{f,i}$, is the actual running time of train f between station i and i', $\bar{r}_{f,i,i'}$ is the recommended running time between those two stations, θ_1 and θ_2 are weights introduced to indicate the relative importance of the two components of the objective function. The term $\max(u_{f,\tilde{i},n},0)$ indicates that regenerative braking is not considered here. The second term in the objective function of the lower level is the deviation with respect to the train schedule specified in the upper level. It is noted that the running time of train f between stations i and i' could be shorter or longer than the predefined running time in the upper level. So even if there are TSRs appearing in the high speed railway line, the train optimal control problem is still feasible in the lower layer.

IV. Hierarchical MPC formulation and solution approach

For the formulated train rescheduling and speed control problem, the train rescheduling problem is executed at a slow time scale in the upper layer, while the speed control problem is implemented in a fast time scale in the lower layer. The hierarchical framework is illustrated in Figure 6, which shows that the original integrated train rescheduling and speed control problem is effectively reduced and decomposed into a two-layer small-scale optimization problem. The reduced upper layer problem is only related to the train rescheduling problem, which can be solved fast when compared to the original integrated problem. The decision variables of the upper layer involve the running/dwell times (i.e., $r_{f,i}$ and $w_{f,i}$) and the arriving/departing order between trains (i.e., $\xi_{f,f',i}$, $\eta_{f,f',i}$, and $\varepsilon_{f,f',i}$), while the departure and arrival times (i.e., $d_{f,i}$ and $a_{f,i}$) of train services are the state variables. In the lower layer, a distributed optimal control structure is formulated for the train speed control problem, where the speed control problem of each train can be solved independently with consideration of the influence of only the preceding train. The decision variables of the lower layer involves the traction/braking forces (i.e., $u_{f,i,n}$), while the unit kinetic energy (i.e., $E_{f,i,n}$) and arrival time (i.e., $a_{f,i,n}$) are the state variables.

A. The MPC problem in the upper layer

The upper layer MPC controller is based on event-triggered scheme instead of time-triggered scheme and it is on a continuous time grid, which is different from the classic MPC scheme based on discrete time models. Hence, we denote the rolling horizon framework via stage k instead of sampling time step k for clarity. The prediction and control horizon T of the
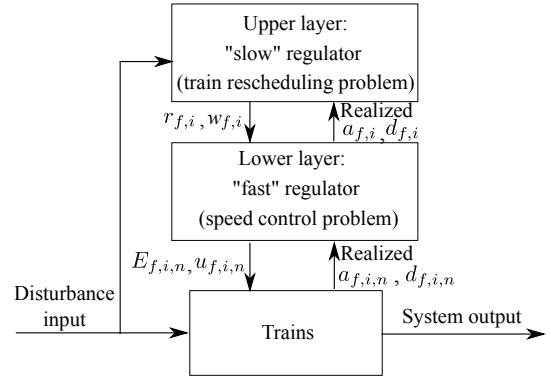

Fig. 6: Hierarchical control structure for train rescheduling and speed control

MPC controller in the upper layer can be chosen as a fixed time interval, e.g., 30 or 45 minutes, according to the practical applications. The number of departure and arrival events in the control horizon T is related to the number and characteristics of train services specified in the planned timetable. We note that the upper layer MPC controller is triggered at a slow time scale when compared with the lower level MPC controller. Specifically, the events that can trigger the MPC controller of the upper level involve the delay events (i.e., the arrivals and departures of trains at stations that are delayed more than a pre-defined threshold) and the TSR events (i.e., the information of TSRs is updated). Hence, the triggered time of stage k is dependent on the arrival/departures events and the TSR events. Once the arrival/departure events or the TSR events occur, the sampling time step k is updated at once. At each sampling time step k, based on the newly available feedback information from the lower level, a real-time optimization problem is formulated to determine the train rescheduling strategy $(a_{f,i},d_{f,i},r_{f,i},\xi_{f,f',i},\eta_{f,f',i})$ for $f \in \mathbb{F}$, $f' \in \mathbb{F}$, $i \in \mathbb{I}_f$, $i \in \mathbb{I}_{f'}$, i.e., the train arrival times, dwell times, running times and the arriving and departing orders over the prediction time horizon T. Since the train rescheduling strategy is based on the indices of train services and stations, we denote the index sets of the relevant train services and stations in predictive time horizon T by F_k and I_k for time step k. The control horizon of the MPC problem in the upper level is the same as the prediction time horizon. Moreover, the control strategy calculated at time step k will be applied to the trains until next event occur, i.e., sampling time step $k+1$.

According to economic criteria for reducing train delays, the optimization problem at the higher layer for time step k over the prediction time horizon T is formulated as follows:

$$\min_{(a^k_{f,i},d^k_{f,i},r^k_{f,i},\xi^k_{f,f',i},\eta^k_{f,f',i})} \sum_{f\in F_k}\sum_{i\in I_k} \frac{|a^k_{f,i}-\bar{a}_{f,i}|}{I^{\text{stop}}_{F_k}} + \frac{|d^k_{f,i}-\bar{d}_{f,i}|}{I^{\text{stop}}_{F_k}} \tag{39}$$
$$\text{s.t.} \quad (2)-(17),$$

for $f \in \mathbb{F}_k$, $f' \in \mathbb{F}_k$, $i \in \mathbb{I}_f$, $i \in \mathbb{I}_{f'}$, where $I^{\text{stop}}_{F_k}$ is the number of scheduled stops for train services in $\mathbb{F}_k$, $a^k_{f,i}$ and $d^k_{f,i}$ are the rescheduled arrival and departure times in the optimization problem at stage k. The optimization problem at the upper level can be transformed into a mixed integer linear programming

(MILP) model. By solving the above optimization problem (39), the controller at the upper layer computes its desired control inputs for the running times of each train between the consecutive stations, which are the reference signals of the train control problem at the lower layer. In the MPC framework, the optimization problem for the train rescheduling is solved online in an event-triggered way with the updated departure/arrival times (i.e., the measurements that feedback from the lower layer) and the updated TSRs information. At the triggered time instant of stage $k+1$, a new optimization problem is formulated again whenever the delay threshold is reached or the TSRs information is updated, which in practice makes the solutions more robust to uncertainty, disturbances, and even model mismatch.

To prove the recursive feasibility of the train rescheduling problem in (39), we first separate the constraints into two sets as follows:

(1) $\mathscr{C}_1(\mathbf{a}^k, \mathbf{d}^k, \mathbf{r}^k, \mathbf{w}^k, \mathbf{Y}^k)$: This is the set of constraints that are related to the running/dwell times, the departure/arrival times, and the binary indicators for the TSR influence, i.e., (2)-(8), at stage k. In particular, we have $\mathbf{a}^k = \{a_{f,i}^k | f \in F_k, i \in I_k\}$ and $\mathbf{d}^k = \{d_{f,i}^k | f \in F_k, i \in I_k\}$. The other sets can be defined in a similar way.'

(2) $\mathscr{C}_2(\xi^k, \eta^k, \varepsilon^k)$: This is the set of constraints that are related to the binary variables for describing the departure/arrival orders among trains, i.e., (9)-(17), at sampling time step k.

We suppose that a feasible solution $\mathscr{S}(\mathbf{a}^{k-1}, \mathbf{d}^{k-1}, \mathbf{r}^{k-1}, \mathbf{w}^{k-1}, \mathbf{Y}^{k-1}, \xi^{k-1}, \eta^{k-1}, \varepsilon^{k-1})$ is available for the train rescheduling problem at time step $k-1$. Many candidate solutions, e.g., $\mathscr{S}(\mathbf{a}^k, \mathbf{d}^k, \mathbf{r}^k, \mathbf{w}^k, \mathbf{Y}^k)$, are recursively feasible to the train rescheduling problem with constraint set $\mathscr{C}_1(\mathbf{a}^k, \mathbf{d}^k, \mathbf{r}^k, \mathbf{w}^k, \mathbf{Y}^k)$ at time step k, because there are no upper limits for the running/dwell times and departure/arrival times. Even though the TSRs would prolong the running times of trains, the problem with constraint set $\mathscr{C}_1(\mathbf{a}^k, \mathbf{d}^k, \mathbf{r}^k, \mathbf{w}^k, \mathbf{Y}^k)$ is always feasible. Moreover, there exists at least one feasible solution that satisfy $\mathscr{C}_1(\mathbf{a}^k, \mathbf{d}^k, \mathbf{r}^k, \mathbf{w}^k, \mathbf{Y}^k)$ and $\mathscr{C}_2(\xi^k, \eta^k, \varepsilon^k)$, i.e., the specified departure/arrival orders in the planned timetable. Hence, the feasibility of the upper layer can be ensured.

B. The MPC problem in the lower layer

In the lower layer, based on the reference of the running times calculated in the higher layer, a speed control problem is implemented in a fast time scale by using the MPC algorithm. With the reference running times of each train obtained from the upper layer, the MPC controllers in the lower level are implemented in a distributed way, where the speed control problem of each train is calculated independently and the influence of the neighboring trains is simplified by only considering the preceding train to reduce the computational complexity significantly. Hence, the communication between the preceding train and the following train is required and the speed profile of the leading train needs to be sent immediately to the following train when the profile is updated. The

distributed MPC problems in the lower level are also event-triggered, where the events are defined as the deviations from the optimized speed profiles are larger than the pre-defined thresholds.

To develop the MPC algorithm for the train control problem at a fast time scale, for each train f, we consider the sampling time step as k_f and the prediction horizon as S_f. The sampling time step k_f is dependent on the departure/arrival events at block section level and the events caused by TSRs. It is worth to note that the prediction horizon of the MPC controller in the lower layer is based on position instead of time. The prediction horizon S_f should cover the distance between the current position of train f and the planned stopping platform. Moreover, the movement authority related information, i.e., the operational status (e.g., position and speed) of the preceding train, is considered in the lower layer problem. During the prediction horizon S_f, the set of the block sections that could be traversed by train f is denoted by N_{k_f, S_f}. The reference running time during the prediction horizon is denoted as r_{f,k_f}, which is calculated by the higher layer. At sampling time step k_f, based on the newly available feedback information, e.g., the movement authority information, a real-time optimization problem is formulated for each train to determine the train operation control strategy $u_{f,i,n}$ as follows:

$$\min_{u_{f,i,n}} \sum_{(\tilde{i},n) \in N_{k_f, S_f}} \theta_1 \max(u_{f,\tilde{i},n}, 0) \Delta s_{\tilde{i},n} + \theta_2 |r_{f,k_f} - \bar{r}_{f,k_f}| \quad (40)$$
$$\text{s.t.} \quad (20) - (36),$$

where the first term is to minimize the energy consumption and the second term is to track the reference running time given by the upper layer. In the lower layer, a set of distributed MPC controllers (one controller for each train) are implemented in the fast time scale. These MPC controllers are implemented as a special case of the shrinking horizon strategy since the prediction horizon in the lower layer is the distance to the next stopping station and it is reduced with the increasing of the sampling time step. The optimization problem given in (40) can be transformed into a MILP problem by approximating the nonlinear terms by piecewise affine functions and applying the transformation properties given by [29] by introducing auxiliary binary-valued and real-valued variables. The resulting MILP problem can then be effectively solved by existing solvers similar as the optimization problem in the upper level. We refer to [28] for more details about the MILP transformation.

The proof of the recursive feasibility of the lower layer is similar as the one for the upper layer. Hence, we explain the reasons for the feasibility briefly for the sake of simplicity. The reference running time specified by the upper layer is considered as a soft constraint in the objective function instead of a hard constraint for the lower level problem. Hence, the updates of the running times in the higher level do not influence the recursive feasibility. The newly triggered TSRs may cause the speed train control problem infeasible because the speed of trains may not be reduced to satisfy the TSRs in some extreme scenarios; however, the dispatchers and the centralized train control system need to check the influence of the TSRs on the operation of trains before setting the TSR

command to avoid in-feasibilities. So the recursive feasibility of the lower layer can also be ensured.

C. The whole hierarchical model predictive control algorithm

The whole hierarchical model predictive control algorithm for the on-line railway delay management and train control is summarized as given in Algorithm 1. In addition, the flow chart of the algorithm proposed in this paper is illustrated in Figure 7.

Algorithm 1 Hierarchical MPC algorithm for on-line railway delay management and train control

---

Step 1. In the upper layer, at each sample step k, obtain the measured feedback information, i.e., departure and arrival times of trains at stations, and the TSR data.

Step 2. Formulate the train rescheduling problem (39) under constrains (2)-(17) for sample step k with given prediction horizon T (i.e., sets of relevant train services and train stations are denoted as F_k and I_k) in a slow time scale.

Step 3. Solve train rescheduling problem (39) for sample step k using existing solvers, e.g., CPLEX, and obtain the updated train departure/arrival times, stop patterns, and running times.

Step 4. Taking running times obtained from the upper layer as references, formulate the lower layer train control problems in a fast time scale and solve them in a distributed way:

  Step 4.1 For each train $f \in F_k$ at the sample step k_f, formulate the train speed control problem (40) under constraints (20)-(36) over the predictive horizon S_f.

  Step 4.2 Solve speed control problem (40) for each train using existing solvers, e.g., CPLEX, and obtain the optimal speed profile and apply the corresponding control input to each train during sample step k_f.

  Step 4.3 In the next step $k_f + 1$, for each train, based on the new measured feedback information of itself and its preceding train, repeat Steps 4.1-4.2 without communications with other trains, until the solution process of the upper layer is triggered, i.e., a train departs from or arrives at a station.

Step 5. Based on the measured feedback information for the train departure times, arrival times, and the TSR data in the next step $k + 1$, repeat Steps 1-4 until the end of the decision horizon.

According to the above hierarchical MPC algorithm, the original complex high speed railway delay management and speed control problem is assigned to different layers, in which different optimal control problems with specific tasks are solved. The hierarchical algorithm effectively reduces the computation burden for the original optimization problem. Specifically, in the higher layer train rescheduling model (that is simpler and more abstract) is presented to predict the long-term behavior of the railway system and to minimize the specified objective function over a long time horizon. In the lower layer, train speed control model (that is more accurate) is adopted to optimize the current control strategies by concentrating on a shorter time horizon. We note that the MPC controller in the upper and lower layers are both event-triggered. Several events would occur during the prediction horizon in general. However, if no events occur during the prediction horizon, the solution process in the upper layer will be triggered at the end of the prediction horizon. For the lower layer, if the prediction horizon is reached, which basically means that a train arrives at the planned station. This would trigger the MPC controllers in the upper and lower layers.
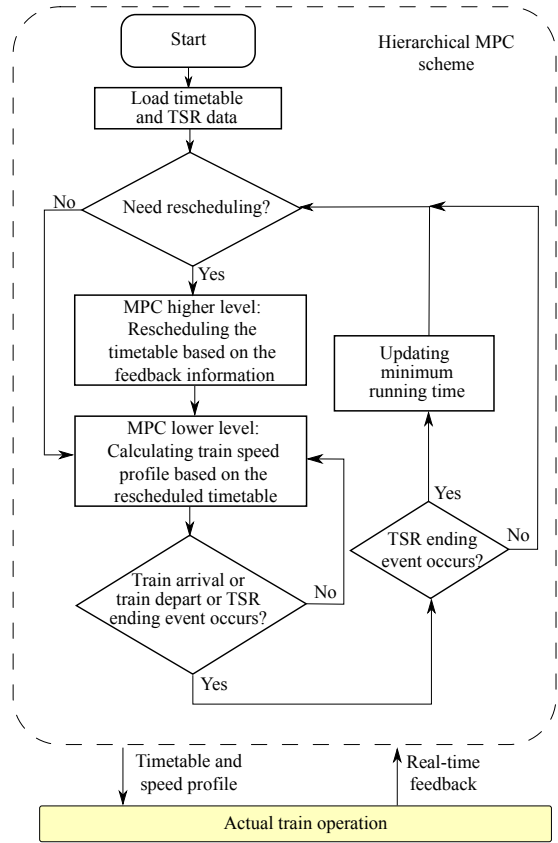

Fig. 7: The flow chart of the hierarchical MPC algorithm

It should be noted that, in the proposed hierarchical model predictive control framework, the train dynamic model is a hybrid discrete event system, and the formulated problems in the upper layer and in the lower layer are both mixed integer linear programming problems, which are non-convex. It becomes more cumbersome to give the rigorous proof of stability for the nonlinear system with the non-convex constraints, which needs to be investigated in our future work. Specifically, the speed control problem for each train is calculated in a distributed manner that only needs the information of its preceding train, not other trains, which greatly improves the computational efficiency compared to the centralized control of all the trains. The modularity of distributed manner benefits that if a new train is added to the system, the algorithm at the group involving the new train is only modified. Moreover, with the hierarchical MPC algorithm, both the upper layer problem and the lower layer problem are formulated as small-scale mixed integer linear programming problems, which can be efficiently solved by the existing solvers, such as CPLEX and GUROBI.

V. CASE STUDY

To illustrate the effectiveness and feasibility of the proposed hierarchical model predictive control scheme, numerical experiments are performed based on the actual data of Beijing-Shanghai high-speed railway. All of these experiments are conducted by using MATLAB 2016b and CPLEX 12.8 on

TABLE III: The position of stations in the Beijing-Shanghai high-speed railway

Station index	Station name	Position (km)
S1	Beijing South	0
S2	Langfang	59
S3	Tianjin South	131
S4	Cangzhou West	219
S5	Dezhou East	327
S6	Jinan West	419
S7	Tai'an	462
S8	Qufu East	533
S9	Tengzhou East	589
S10	Zaozhuang	625
S11	Xuzhou East	688
S12	Suzhou East	767
S13	Bengbu South	844
S14	Dingyuan	897
S15	Chuzhou	959
S16	Nanjing	1018
S17	Zhenjiang South	1087
S18	Danyang North	1112
S19	Changzhou North	1144
S20	Wuxi East	1201
S21	Suzhou North	1227
S22	Kunshan South	1259
S23	Shanghai Hongqiao	1302

TABLE IV: The TSR information

Index	Time period	Location	Speed restriction (km/h)
1	[10:00, 10:30]	[S8, S9]	160
2	[10:30, 11:00]	[S4+1km, S4+81km]	200
3	[10:30, 11:20]	[S2, S2+41km]	150
4	[11:00, 11:30]	[S21+8km, S21+23km]	160

a personal computer (Windows 10 operating system, Intel(R) Core i7-7700HQ CPU @2.80GHz, 8.0 GB RAM).

A. The set-up of the case study

The Beijing-Shanghai high-speed railway is one of the busiest high-speed railway lines in China; its the total length is 1318 kilometers. There are 23 stations on this high-speed railway line and the exact locations of these stations are listed in Table III. For our case study, the planned (or original) timetable is based on the actual timetable of the Beijing-Shanghai high-speed railway on a day of 2019. The time period considered in this paper is from 10:00 am to 12:00 am and the number of involved train services is 33 for this time period. For the train services that have started running before 10:00 am, we assume that they are operated according to the planned timetable before 10:00, i.e., no delays at the beginning of the considered time period. However, the disturbances is considered for the operation of trains during the time period. It is noted that the planned timetable of the Beijing-Shanghai high-speed railway is a non-cyclic timetable. The temporary speed restrictions (TSRs) that are considered in this case study are given in Table IV, where four TSRs are considered at four different locations of the line. The activation periods of these four TSRs are between 10:00 am and 11:30 am, while different TSRs appear at different time periods.

The maximum running speed of the considered trains is 300 km/h. The minimum arrival and departure headway of two consecutive trains are chosen as 3 minutes at stations. Moreover, the minimum running times of trains between stations is set as 90% of the planned running times as specified in the planned timetable. The values of θ_1 and θ_2 in objective function (40) are taken as $1 \cdot 10^{-9}$ and 30, respectively. The disturbances of running times and dwell times, i.e., $\sigma_{f,i}^{r}$ and $\sigma_{f,i}^{w}$, in the higher level obey uniform distribution in our case study. More specifically, the disturbances of running times range from 30 s to 60 s and the disturbances of dwelling times

at a stations range from 0 s to 30 s. In addition, disturbances that also obey uniform distributions are introduced for the lower level to indicate the tracking errors of the optimized speed profiles, which range from 0 s to 90 s.

B. Comparison of the proposed HMPC strategy with other policies

To evaluate the effectiveness of the hierarchical control scheme, we compare it with two widely accepted train regulation policies, i.e., the FSFS (first-scheduled-first-served) policy and the FCFS (first-come-first-served) policy. The detailed settings for those three strategies are given as follows:

- Hierarchical model predictive control (HMPC) strategy: The control horizon and prediction horizon in the upper layer of the HMPC framework are both set as 30 minutes. The control horizon and prediction horizon in the lower layer are usually taken as the left running times to the next stop, which are much shorter when compared to those of the upper layer. The calculations in the framework are event-triggered. Particularly, the events that trigger the computations in the upper layer are train delay events, TRS events, etc. Specifically, when a train has a delay of more than 1.5 minutes at stations, the train schedule in the upper layer and speed profiles in the lower layer need to be updated. Moreover, when the error between the operation speed of trains and the planned speed profile is bigger than a predefined threshold, i.e., 2 km/h, the speed profile of a particular train needs to be optimized again. The state-of-the-art model predictive control solutions are calculated for the upper hierarchical layer and the lower hierarchical layer, respectively.

- FSFS strategy: The order of trains that depart from stations in practice will follow the order specified in the planned timetable. If possible, the dwell times and running times of trains should stay the same as the ones in the planned timetable; however, they will be affected by the running and dwell time disturbances. Moreover, if the operation of trains is affected by the TSRs, the minimum running time between stations should change accordingly. For this study, the minimum running times under the TSRs are calculated via the detailed train characteristics model. Furthermore, when the train schedule is updated according to the FSFS policy, the speed curves of the trains and the corresponding energy consumption are recalculated.

- FCFS strategy: The order of trains follows the actual arrival order at stations. The running times and dwell

TABLE V: The performance comparison of the proposed HMPC strategy, FSFS policy, and FCFS policy

Performance index	HMPC	FSFS	FCFS
Delay (min)	278.66	929.16	1005.16
Average energy consumption (kW·h/km)	14.31	12.12	12.14

times of trains in the FCFS strategy are obtained similarly as what we have for the FSFS strategy.

In the lower level, extra speed limitation constraints may be added to the train control model according to the number of free block sections between the current train and the preceding train via (37).

In this experiment, the higher level of the HMPC framework is triggered 13 times during the considered time period and the computation time of these calculations ranges from 20 s to 26 s for the optimization problem in the higher level. The MPC controllers in the lower level are triggered 415 times during the considered time period and the average computation time is around 16 s. We note that the triggered frequency of different trains are different, which depends on the operation status of trains and the TSRs. For example, the controllers of train 16 and 17 are triggered 12 and 22 times, respectively. Specifically, the numbers of the constraints, the real-valued variables, and the integer-valued variables are about 10000, 5000, and 45000, respectively, in the upper layer problem. In the lower layer problems, the number of the constraints ranges from 10000 to 15000; the number of the real-valued variables ranges from 2000 to 3000; and the numbers of the integer-valued variables ranges from 18000 to 27000. Based on the computation times, the higher layer of the HMPC framework can satisfy the real-time requirement in general, while more efficient algorithms need to be investigated for the lower layer . In addition, the computation times of the FSFS and FCFS are much smaller than that of the higher level of the HMPC framework, and they are equal to 0.42 s and 0.39 s, respectively.

The performance comparison of these three strategies is given in Table V, where the departure/arrival time deviations (mostly delays) and the average energy consumption per kilometer are reported. As can be observed from Table V, the proposed HMPC strategy reduces the delay by 70.01% and 72.28% when compared with the FSFS and FCFS policies. However, the delay reduction of the HMPC strategy is achieved at the cost of a higher energy consumption. Particularly, the average energy consumption per kilometer of the HMPC strategy increases by 18.07% and 17.87% when compared with the FSFS and FCFS policies. In addition, the rescheduled timetable obtained by the HMPC is illustrated in Figure 8. The detailed comparison among the rescheduled solutions computed by the HMPC, FSFS, and FCFS strategies are given in Figures 9 and 10. In Figure 8, the running times of trains prolonged if their operations are affected by the TSRs. Moreover, it can be observed from Figure 8 that the operations of some trains that are not affected by TSRs may slightly deviate from the planned timetable, e.g., trains 5 and 30. This

is because the upper layer of the HMPC approach involves the estimations of the running and dwell time disturbances, which affect the actual arrival and departure times of trains at stations.

In Figure 9, the arrival times of trains 8, 13, 16, and 19 obtained by the HMPC strategy are earlier than those obtained by FSFS and FCFS strategies, which means less delays. In addition, we note that trains 16 and 19 runs faster before the TSRs, i.e., between stations S1 and S2, to reduce delays for the HMPC approach. Moreover, the speed profiles of train 16 obtain by these three strategies are given in Figure 11, where the speed profiles of the HMPC strategy is much higher than the FSFS and FCFS strategies. Train 16 plans to depart from station S2 at 10:58 and arrive at station S3 at 11:16, so it is influenced by TSR 3 of Table IV. The running times between stations, the delays, and the energy consumption of the three strategies are reported in Table VI for train 16. Due to the TSR between station S2 and station S3 (as can be seen from Figure 11), the running time of train 16 between S1 and S3 is longer than the planned running time, i.e., 36 minutes, in the original timetable. As can be seen from Table VI, the running times obtained by the HMPC strategy are shorter than those of the FSFS and FCFS strategies. However, the energy consumption of the HMPC strategy is much higher than that of the FSFS and FCFS strategies. The faster running of trains results higher energy consumption. In Figure 10, the departure order of train 12 and 22 obtained by different strategies are different. Specifically, train 12 departs from station S5 earlier than train 22 in the HMPC and FSFS strategies, while the departure order is different in the FCFS strategy.
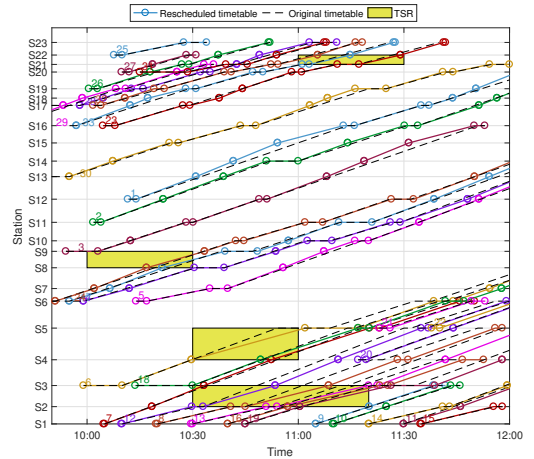

Fig. 8: The train rescheduled solutions obtained by the HMPC strategy

It is worth to note that the delay recovery is much more important than the energy consumption in general during disruptions. So the higher energy consumption of the HMPC approach is acceptable in practical applications. Moreover, the trade-off between the total delay and the energy consumption is investigated by adjusting the weights in the lower layer problem, as shown in Figure 12. We note that with the increase of the relative importance of the energy consumption, the energy consumption of the HMPC approach decreases.
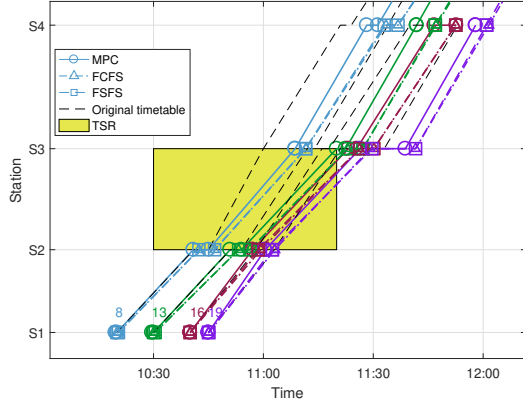
Fig. 9: Train rescheduling solutions of trains 8, 13, 16, and 19 obtained by the HMPC, FSFS, and FCFS strategies
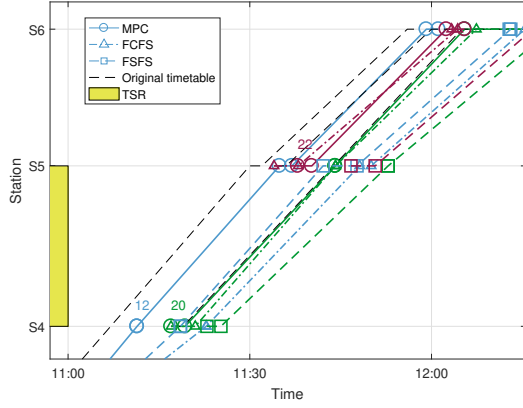

Fig. 10: Train rescheduling solutions of trains 8, 13, 16, and 19 obtained by the HMPC, FSFS, and FCFS strategies
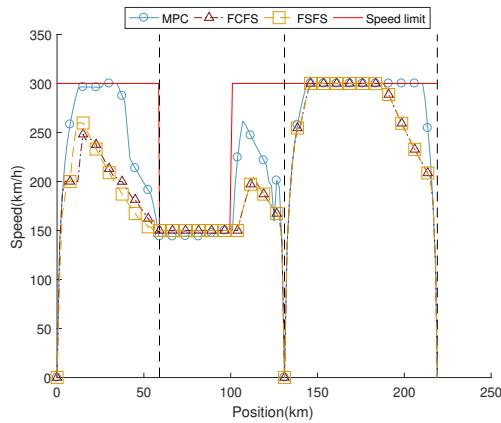

Fig. 11: Speed profiles of train 16 obtained by the HMPC, FSFS, and FCFS strategies

TABLE VI: Running times and energy consumption of train 16 obtained by the HMPC, FSFS, and FCFS strategies

Strategy	HMPC		FSFS		FCFS	
Station sections	S1-S3	S3-S4	S1-S3	S3-S4	S1-S3	S3-S4
Running time (min)	43.15	21.20	46.29	22.34	46.29	22.34
Delay (min)	7.15	6.35	10.30	12.51	10.30	12.51
Energy consumption (kW·h)	1174.55	1561.67	843.99	1251.68	844.40	1251.66

Specifically, when the delays obtained by the HMPC approach is larger than 600 minutes, the resulting energy consumption is smaller than the FCFS and FSFS approaches. It can be concluded that the HMPC approach could reduce the delay and energy consumption at the same time.
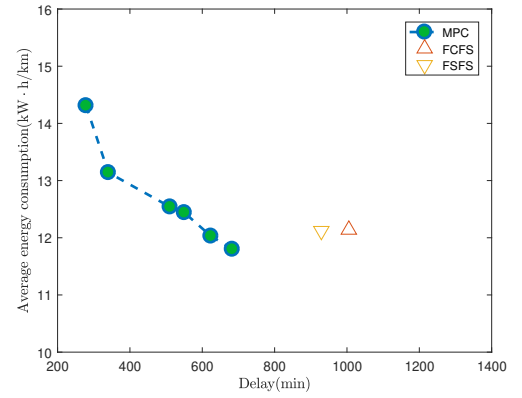

Fig. 12: The trade-off between the delay management and the energy consumption

C. Robustness performance of the proposed HMPC approach

TABLE VII: Robustness comparison of the tested 100 scenarios for the proposed HMPC approach (SD and CV stand for the standard deviation and the coefficient of variation)

Indicators	Mean	SD	CV
Total disturbance (min)	114.06	8.09	7.10%
Total delay (min)	286.58	6.59	2.30%
Average energy consumption (kW·h/km)	14.30	0.068	0.47%

In the actual operation of high-speed railways, disturbances occur due to various reasons, which makes the disturbances uncertain and unpredictable. So there is a high robustness requirement for a rescheduling method. The results of a good train rescheduling approach should not change largely when the disturbances vary. To analyze the robustness of the proposed HMPC approach, a series of numerical experiments is performed. We generate different disturbances in the higher

level randomly according to the uniform distributions for different scenarios, where the disturbance at each station ranges from 0.5 minute to 1 minute. The considered time period, original timetable, and the TSRs information are the same as those of the previous experiment. The robustness analysis results of the 100 disturbed scenarios are illustrated in Table VII, where the average values of the total disturbance, the total delay, the energy consumption per kilometer, the mean value, the standard deviation, and the coefficient of variation are listed. Specifically, the coefficient of variation is defined as the percentage between the standard deviation and the mean value. The total disturbance ranges from 81.68 minutes to 122.73 minutes, while the total delay ranges from 273.47 minutes to 301.15 minutes and the average energy consumption per kilometer ranges from 14.11 kW·h to 14.49 kW ·h. We note that the coefficient of variation of the total disturbance, i.e., 7.10%, is much larger when compared with those of the total delay and average energy consumption. Hence, the the proposed HMPC approach has good robustness to disturbance in general.
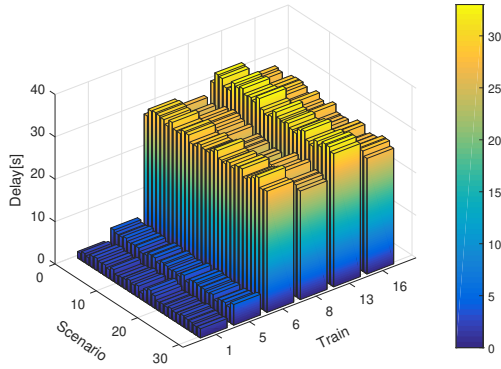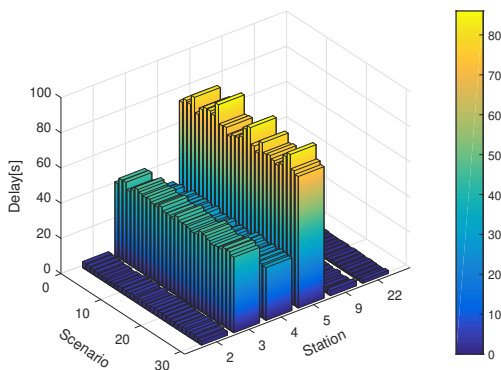

Fig. 13: Delays of several trains in 30 scenarios

Fig. 14: Delays of several stations in 30 scenarios

More specifically, the total delay of several trains and the total delay at several stations are shown in Figure 13 and Figure 14 for 30 disturbed scenarios , where the z-axis represents the total delay. It can be observed that there is no obvious change for different scenarios. In Figure 13, trains 1

and 5 have much less delay when compared with the other trains, because they are not influenced by TSRs. For train 13, the average total delay for the tested 100 scenarios is 31.28 minutes and the standard deviation is 0.83 minutes, which means that the fluctuation is relatively small. In Figure 14, stations 2, 9, and 22 have much smaller delays when compared with other stations. The reasons are given as follows: (1) there is no TSR between stations 1 and 2; (2) a few trains are influenced by the TSR between stations 8 and 9 and the TSR between stations 21 to 22. For station 3, the average delay of the tested 100 scenarios is 43.49 minutes and the standard deviation is 1.13 minutes. Therefore, it can be concluded that the proposed HMPC approach is relatively robust to the fluctuations of disturbances.

VI. CONCLUSIONS AND FUTURE RESEARCH

In this paper, we have tackled the on-line high-speed railway delay management and train control problem under a dynamic operation environment. A practical and effective hierarchical model predictive control (MPC) framework has been proposed, where train schedules and speed profiles are updated based on the real-time updated information. Moreover, different updating frequencies are adopted for the train rescheduling in the upper control level and the speed profile optimization in the lower control level. The optimization problems in the upper layer and lower layer are both formulated as mixed integer linear programming problems, which can be solved efficiently with existing solvers. The effectiveness and robustness of the proposed two-level hierarchical control framework has been analyzed through several numerical experiments based on the data from the Beijing-Shanghai high-speed railway. When compared with the widely used strategies, the presented hierarchical MPC framework can yield smaller train delays at the cost of an acceptable increase in the energy consumption. Moreover, the hierarchical MPC framework has good robustness performance for different disturbances.

To investigate the trade-off between delay management and energy consumption better, an interesting research direction is to integrate a simple dynamical train model in the current model of the upper layer. Hence, the energy consumption is also included in the performance index of the upper layer, which would improve the performance and consistency of the proposed hierarchical MPC framework. We note that the computation efficiency of the lower level controllers is not sufficient for the real-time application. In our future work, more efficient algorithms should be designed to update the train speed profiles. In addition, the robustness of the proposed hierarchical MPC approach is important for practical applications. We would like to investigate on a robust MPC scheme and evaluate its performance in our future work. Another promising future work direction is to extend the hierarchical optimization problem for on-line delay management and train control in a large-scale high-speed railway network, where a distributed optimization framework could be introduced for the train rescheduling and control problem to reduce the computational complexity.

TABLE VIII: Numbers of variables and constraints in the train rescheduling problem

Variables or constraints	Total number at most
Variable $a_{f,i}$	$\sum_{f\in\mathbb{F}}|\mathbb{I}_f|$
Variable $d_{f,i}$	$\sum_{f\in\mathbb{F}}|\mathbb{I}_f|$
Variable $r_{f,i}$	$\sum_{f\in\mathbb{F}}(|\mathbb{I}_f|-1)$
Variable $w_{f,i}$	$\sum_{f\in\mathbb{F}}|\mathbb{I}_f|$
Binary variable $Y_{f,i,\ell}$	$\sum_{\ell\in\mathbb{L}}\sum_{i,i+1\in\mathbb{I}_\ell}|\mathbb{F}_{i,i+1}|$
Binary variable $\xi_{f,f',i}$	$\sum_{i\in\mathbb{I}}|\mathbb{F}_i|(|\mathbb{F}_i|-1)$
Binary variable $\eta_{f,f',i}$	$\sum_{i\in\mathbb{I}}|\mathbb{F}_i|(|\mathbb{F}_i|-1)$
Binary variable $\varepsilon_{f,f',i}$	$\sum_{i\in\mathbb{I}}|\mathbb{F}_i|(|\mathbb{F}_i|-1)$
Auxiliary variables	$2\sum_{\ell\in\mathbb{L}}\sum_{i,i+1\in\mathbb{I}_\ell}|\mathbb{F}_{i,i+1}|$
Departure/arrival constraints	$2\sum_{f\in\mathbb{F}}(|\mathbb{I}_f|-1)$ $+3\sum_{f\in\mathbb{F}}|\mathbb{I}_f|$ $+8\sum_{\ell\in\mathbb{L}}\sum_{i,i+1\in\mathbb{I}_\ell}|\mathbb{F}_{i,i+1}|$
Headway and ordering constraints	$7\sum_{i\in\mathbb{I}}|\mathbb{F}_i|(|\mathbb{F}_i|-1)$ $+\sum_{i,i+1\in\mathbb{I}}|\mathbb{F}_{i,i+1}|(|\mathbb{F}_{i,i+1}|-1)$
Station capacity constraints	$2\sum_{i\in\mathbb{I}}|\mathbb{F}_i|(|\mathbb{F}_i|-1)$ $+\sum_{f\in\mathbb{F}}|\mathbb{I}_f|$

Appendix

A. Number of variables and constraints in the Hierarchical MPC approach

The numbers of variables and constraints of the train rescheduling problem in the upper layer are given in Table VIII, where $|\mathbb{F}|, |\mathbb{I}|, |\mathbb{I}_f|, |\mathbb{F}_i|$ and $|\mathbb{F}_{i,i+1}|$are the numbers of elements in $\mathbb{F}, \mathbb{I}, \mathbb{I}_f, \mathbb{F}_i$ and $\mathbb{F}_{i,i+1}$, respectively. Similar analysis can be done for the lower optimization problem.

References

[1] S. Narayanaswami and N. Rangaraj, "Scheduling and rescheduling of railway operations: A review and expository analysis," *Technology Operation Management*, vol. 2, no. 2, pp. 102–122, 2011.

[2] V. Cacchiani, D. Huisman, M. Kidd, L. Kroon, P. Toth, L. Veelenturf, and J. Wagenaar, "An overview of recovery models and algorithms for real-time railway rescheduling," *Transportation Research Part B: Methodological*, vol. 63, pp. 15–37, 2014.

[3] F. Corman and L. Meng, "A review of online dynamic models and algorithms for railway traffic management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1274–1284, 2014.

[4] X. Yang, X. Li, B. Ning, and T. Tang, "A survey on energy-efficient train operation for urban rail transit," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 2–13, 2015.

[5] G. M. Scheepmaker, R. M. Goverde, and L. G. Kroon, "Review of energy-efficient train control and timetabling," *European Journal of Operational Research*, vol. 257, no. 2, pp. 355–376, 2017.

[6] J. Yin, T. Tang, L. Yang, J. Xun, Y. Huang, and Z. Gao, "Research and development of automatic train operation for railway transportation systems: A survey," *Transportation Research Part C: Emerging Technologies*, vol. 85, pp. 548–572, 2017.

[7] W. Fang, S. Yang, and X. Yao, "A survey on problem models and solution approaches to rescheduling in railway networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 2997–3016, 2015.

[8] T. Albrecht, "The influence of anticipating train driving on the dispatching process in railway conflict situations," *Networks and Spatial Economics*, vol. 9, no. 1, pp. 85–101, 2009.

[9] A. D'Ariano, T. Albrecht, J. Allan, C. Brebbia, and A. Rumsey, "Running time re-optimization during real-time timetable perturbations," *Timetable Planning and Information Quality*, vol. 1, pp. 147–156, 2010.

[10] F. Corman, A. D'Ariano, D. Pacciarelli, and M. Pranzo, "Evaluation of green wave policy in real-time railway traffic management," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 6, pp. 607–616, 2009.

[11] G. Caimi, M. Fuchsberger, M. Laumanns, and M. Lüthi, "A model predictive control approach for discrete-time rescheduling in complex central railway station areas," *Computers & Operations Research*, vol. 39, no. 11, pp. 2578–2593, 2012.

[12] M. Mazzarello and E. Ottaviani, "A traffic management system for real-time traffic optimisation in railways," *Transportation Research Part B: Methodological*, vol. 41, no. 2, pp. 246–274, 2007.

[13] M. Lüthi, *Improving the Efficiency of Heavily Used Railway Networks through Integrated Real-Time Rescheduling*. ETH Zürich, Switzerland: Ph.D. thesis, 2009.

[14] P. Wang and R. M. Goverde, "Multiple-phase train trajectory optimization with signalling and operational constraints," *Transportation Research Part C: Emerging Technologies*, vol. 69, pp. 255–275, 2016.

[15] J. Yin, T. Tang, L. Yang, Z. Gao, and B. Ran, "Energy-efficient metro train rescheduling with uncertain time-variant passenger demands: An approximate dynamic programming approach," *Transportation Research Part B: Methodological*, vol. 91, pp. 178–210, 2016.

[16] P. Xu, F. Corman, Q. Peng, and X. Luan, "A train rescheduling model integrating speed management during disruptions of high-speed traffic under a quasi-moving block system," *Transportation Research Part B: Methodological*, vol. 104, pp. 638–666, 2017.

[17] X. Rao, M. Montigel, and U. Weidmann, "A new rail optimisation model by integration of traffic management and train automation," *Transportation Research Part C: Emerging Technologies*, vol. 71, pp. 382–405, 2016.

[18] X. Luan, Y. Wang, B. De Schutter, L. Meng, G. Lodewijks, and F. Corman, "Integration of real-time traffic management and train control for rail networks-part 1: Optimization problems and solution approaches," *Transportation Research Part B: Methodological*, vol. 115, pp. 41–71, 2018.

[19] ——, "Integration of real-time traffic management and train control for rail networks-part 2: Extensions towards energy-efficient train operations," *Transportation Research Part B: Methodological*, vol. 115, pp. 72–94, 2018.

[20] N. Geroliminis, J. Haddad, and M. Ramezani, "Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 348–359, 2012.

[21] M. Yildirimoglu, I. I. Sirmatel, and N. Geroliminis, "Hierarchical control of heterogeneous large-scale urban road networks via path assignment and regional route guidance," *Transportation Research Part B: Methodological*, vol. 118, pp. 106–123, 2018.

[22] Y. Zhou, M. Wang, and S. Ahn, "Distributed model predictive control approach for cooperative car-following with guaranteed local and string stability," *Transportation Research Part B: Methodological*, vol. 128, pp. 69 – 86, 2019.

[23] A. Duret, M. Wang, and A. Ladino, "A hierarchical approach for splitting truck platoons near network discontinuities," *Transportation Research Procedia*, vol. 38, pp. 627–646, 2019.

[24] R. Liu and I. Golovicher, "Energy-efficient operation of rail vehicles," *Transportation Research Part A: Policy and Practice*, vol. 37, no. 10, pp. 917–931, 2003.

[25] R. Franke, P. Terwiesch, and M. Meyer, "An algorithm for the optimal control of the driving of trains," in *Proceedings of the 39th IEEE Conference on Decision and Control*, Sydney, Australia, 2003, pp. 2123–2128.

[26] E. Khmelnitsky, "On an optimal control problem of train operation," *IEEE Transactions on Automatic Control*, vol. 45, no. 7, pp. 1257–1266, 2000.

[27] P. Howlett, "The optimal control of a train," *Annals of Operations Research*, vol. 98, no. 1-4, pp. 65–87, 2000.

[28] Y. Wang, B. De Schutter, T. van den Boom, and B. Ning, "Optimal trajectory planning for trains – A pseudospectral method and a mixed integer linear programming approach," *Transportation Research Part C: Emerging Technologies*, vol. 29, pp. 97–114, 2013.

[29] H. P. Williams, *Model building in mathematical programming*. John Wiley & Sons, 2013.